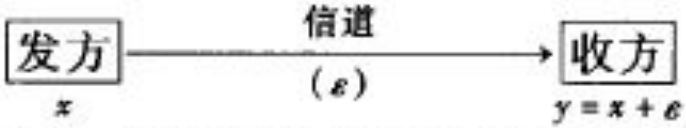
Code 解题报告

AHdoc

问题回顾

通信系统的最简单模型可以表示成：



发送方把信息 x 传给接收方，但是在信道传输过程中出现错误（或干扰）ε，从而收方收到的是 $y=x+\varepsilon$. 我们希望有一种办法，使得收方在得到 y 之后，有能力检查是否有错（$\varepsilon=0$）？并且在有错时（$\varepsilon\neq0$），有能力决定出错误 ε，从而得到所传送的正确信息 $x=y-\varepsilon$（纠错）.

问题回顾

例如我们需要传送 16 个信息给对方，通常把它们表示成$\mathbb{F}_2^4$ 中 16 个元素 (0000)，(1000)，(0100)，(1100)，…，(1111)．如果发方希望把 $x=(0100)$（代表的信息）传给对方，而信息发生错误 $\varepsilon=(0010)$，即在第 3 位出错，则收方得到 $y=x+\varepsilon=(0110)$．这个向量也是有意义的，它也代表某种信息．收方在得到 y 之后，无法判断是否有错．所以，这种编码方式没有任何检错和纠错能力．

为了使通信系统有检错和纠错能力，需要把表示信息的长度加大．

历史上最早被使用的方法

例 1（重复码） 把每个信息$(a_0a_1a_2a_3)\in \mathbb{F}_2^4$重复三次而传送成

$$(a_0a_1a_2a_3a_0a_1a_2a_3a_0a_1a_2a_3)\in \mathbb{F}_2^{12}.$$

所以$\mathbb{F}_2^{12}$中只有2^4个向量才表示有意义的信息，它们形成$\mathbb{F}_2^{12}$的一个子集合

$$C=\{(a_0a_1a_2a_3a_0a_1a_2a_3a_0a_1a_2a_3)\ :\ a_i\in \mathbb{F}_2\}$$

叫做一个纠错码，C中向量叫做码字，而$\mathbb{F}_2^{12}$中其余$2^{12}-2^4$个向量不代表信息（没有意义）.

现在把码字$x=(010001000100)$传给对方. 如果信道只产生一个错位，例如$\varepsilon=(100000000000)$，则收到的$y=x+\varepsilon=(110001000100)$不是码字（没有意义），对方便可发现有错. 进而，把收到的y分成三段：1100，0100，0100.

利用这样的方法我们可以得到 算法一, 期望得分: >= 10.

历史上最早被使用的方法

例 2（奇偶校验码） 将$\mathbb{F}_2^4 = \{(a_0a_1a_2a_3) : a_i \in \mathbb{F}_2\}$中 16 个信息编成$\mathbb{F}_2^5$中 16 元子集合，把$(a_0a_1a_2a_3)$改成用$(a_0a_1a_2a_3a_4)$，其中$a_4 = a_0 + a_1 + a_2 + a_3 \in \mathbb{F}_2$. 例如：(0110)改成(01100)，而(0100)改成(01001). 所以增加的a_4为 0 或 1，使得$(a_0a_1a_2a_3a_4)$当中有偶数个分量为 1. 从而纠错码为：

$$C = \{(a_0a_1a_2a_3a_4) \in \mathbb{F}_2^5 : a_0 + a_1 + a_2 + a_3 + a_4 = 0\}.$$

如果发生 1 位错误，则码字变成分量具有奇数个 1 的向量，它不是码字. 所以 C 可检查 1 位错误，易知它没有任何纠错能力. 这个纠错码的效率为$\frac{4}{5}$.

反问题

定义 有限域$\mathbb{F}_q$上n维向量空间$\mathbb{F}_q^n$的每个(非空)子集C都叫做是一个q元纠错码,C中向量$c=(c_1,c_2,\cdots,c_n)$叫做码字.n叫该码的码长,$K=|C|$为码字个数,$k=\log_q K$叫码的信息位数(它的意义如下:若不考虑纠错问题,K个信息用q元域中元素来表示每位,只需k位即可).而$\frac{k}{n}$叫码C的传输效率$(0<\frac{k}{n}\leqslant 1)$.

除了n和K(或k)之外,码C的另一重要参数是最小距离,它反映该码C的纠错能力.

定义 对于$v=(v_1,v_2,\cdots,v_n)$,$u=(u_1,u_2,\cdots,u_n)\in\mathbb{F}_q^n$,用$w_H(v)$表示非零分量$v_i(1\leqslant i\leqslant n)$的个数,叫做向量$v$的汉明重量.而$d_H(u,v)=w_H(u-v)$叫做向量$u$和$v$之间的汉明距离.即

$$w_H(v)=\#\{i\ :\ 1\leqslant i\leqslant n, v_i\neq 0\},$$

$$d_H(u,v)=\#\{i\ :\ 1\leqslant i\leqslant n, u_i\neq v_i\}\quad (u\text{和}v\text{的相异位个数}).$$

以后$w_H(v)$和$d_H(u,v)$简记为$w(v)$和$d(u,v)$.

纠错码C的最小距离d

下面的结果是纠错码理论的出发点，表明最小距离这个概念确实反映出码的纠错能力.

定理 设纠错码 C 的最小距离为 d，则此码可检查 $\leqslant d-1$ 位错误，也可纠正 $\leqslant \left[\frac{d-1}{2}\right]$ 位错误.

对给定的 F_q，q 元纠错码 C 有三个基本参数：码长 n，信息位数 k（或 $K=|C|=q^k$）和最小距离 d. 这个 q 元码也可表示成 $[n,k,d]$ 或者 (n,K,d).

算法二

- 枚举n
  - 在q^n中搜索到q^k个元素,组成子空间C
  - C的距离大于等于给定值
- 满足条件的最小n就是可行的答案

- 期望得分: >=20

线性码

定义 $\mathbb{F}_q^n$ 的每个$\mathbb{F}_q$ 上的线性子空间 C 叫做 q 元线性码. 换句话说，$\mathbb{F}_q^n$ 的子集合C 叫做 q 元线性码，是指：

若 $c, c' \in C, \alpha, \alpha' \in \mathbb{F}_q$，则 $\alpha c + \alpha' c' \in C$.

引理 对于线性码 C，

$$d(C) = \min\{w(c) : 0 \neq c \in C\}.$$

生成阵G

对于线性码 C,我们可以充分地利用线性代数作为数学工具. 作为 $\mathbb{F}_q^n$ 的 k 维线性子空间,C 可取一组 $\mathbb{F}_q$ 基

$$u_i = (a_{i1}, a_{i2}, \cdots, a_{in}) \quad (1 \leqslant i \leqslant k),$$

其中 $a_{ij} \in \mathbb{F}_q$. 于是每个码字可唯一表示成

$$\begin{aligned} c &= b_1 u_1 + b_2 u_2 + \cdots + b_k u_k \quad (b_i \in \mathbb{F}_q) \\ &= (b_1, b_2, \cdots, b_k) G, \end{aligned}$$

其中 G 是 $\mathbb{F}_q$ 上秩为 k 的 $k \times n$ (k 行 n 列)的矩阵

$$G = \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix}.$$

G 叫做线性码 C 的一个生成阵.

校验阵

q 元线性码 $C=[n,k]$ 是 $\mathbb{F}_q^n$ 的一个 k 维向量子空间. 由线性代数知 C 必是某个线性齐次方程组

$$\begin{cases} b_{11}x_1+b_{12}x_2+\cdots+b_{1n}x_n=0 \\ b_{21}x_1+b_{22}x_2+\cdots+b_{2n}x_n=0 \\ \cdots\cdots\cdots \\ b_{n-k,1}x_1+b_{n-k,2}x_2+\cdots+b_{n-k,n}x_n=0, \end{cases} \quad \text{即 } H\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}=\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

的全部解,其中

$$H=(b_{ij})_{1\leqslant i\leqslant n-k,1\leqslant j\leqslant n}$$

是 $\mathbb{F}_q$ 上 $(n-k)\times n$ ($n-k$ 行 n 列)的矩阵,并且 H 的秩为 $n-k$. H 叫线性码的一个校验阵,因为由定义可知,对于每个 $v\in\mathbb{F}_q^n$,

$$v\in C\Leftrightarrow vH^{\mathrm{T}}=0\ (\text{长为 } n-k \text{ 的零向量}),$$

一个线性码的例子

例 （奇偶校验码）

$$C=\{(b_1,b_2,\cdots,b_n)\in \mathbb{F}_q^n \ : \ b_1+b_2+\cdots+b_n=0\},$$

这是 n 元线性码，$(1,0,\cdots,0,-1),(0,1,0,\cdots,0,-1),\cdots,(0,\cdots,0,1,-1)$ 是它的一组基，从而生成阵可取为

$$G=\begin{bmatrix} & & -1 \\ & I_{n-1} & \vdots \\ & & -1 \end{bmatrix}.$$

码的参数为 $[n,n-1,2]$（易知 C 的最小距离为 2）. 当 $q=2$ 时，此码即是 $\mathbb{F}_2^n$ 中汉明重量为偶数的全部向量组成的线性码

校验阵与C的最小距离

引理 6.2.3 设 C 是参数 $[n,k]$ 的 q 元线性码，$H=[u_1,u_2,\cdots,u_n]$ 是 C 的一个校验阵. 如果 $u_1,u_2,\cdots,u_n$ 当中任意 $d-1$ 个列向量均 $\mathbb{F}_q$ 线性无关，并且存在其中 d 个列向量是 $\mathbb{F}_q$ 线性相关的，则 d 即是线性码 C 的最小距离.

算法三_汉明码

$\mathbb{F}_q$ 上长为 m 的所有非零向量 $0 \neq v = (v_1, v_2, \cdots, v_m) \in \mathbb{F}_q^m$ 共 $q^m - 1$ 个($m \geqslant 2$),其中两个非零向量 v 和 u 线性相关,当且仅当存在 $\alpha \in \mathbb{F}_q^*$ 使得 $v = \alpha u$,这样两个非零向量称作是等价的.彼此等价的非零向量形成一个等价类.由于每个等价类中有 $q-1$ 个非零向量($\alpha v \ : \ \alpha \in \mathbb{F}_q^*$),所以共有 $n = \dfrac{q^m - 1}{q - 1}$ 个等价类.我们从每个等价类中取出一个代表向量,n 个代表向量(均表成列向量)排成 $\mathbb{F}_q$ 上一个 $m \times n$ 的矩阵

$$H = [u_1, u_2, \cdots, u_n].$$

这个矩阵的秩为 m,因为列向量中包含 m 个线性无关的 $\begin{bmatrix} \alpha_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \alpha_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha_m \end{bmatrix}$,其中 $\alpha_1, \alpha_2, \cdots, \alpha_m \in \mathbb{F}_q^*$.

算法三_汉明码

这个码的码长为 $n=\frac{q^m-1}{q-1}$，信息位数为 $n-m=\frac{q^m-1}{q-1}-m$. 由于 H 中诸列均是非零向量，并且 $u_j(1\leqslant j\leqslant n)$ 属于不同等价类，所以 H 中每一列和每两不同列都线性无关. 但是 $u_1+u_2(\neq 0)$ 必与 H 的某一列等价，所以 H 中有 3 个不同列线性相关. 这表明上述 q 元汉明码的参数为 $[n,k,d]=\left[\frac{q^m-1}{q-1},\frac{q^m-1}{q-1}-m,3\right]$.

只能解决d=3的情况,期望得分: >=20

算法四_多项式码

- 考虑n<=q的情况:

设 $a_1, a_2, \cdots, a_n$ 是 $\mathbf{F}_q$ 中 n 个不同元素

$$G=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & & \vdots \\ a_1^{k-1} & a_2^{k-1} & \cdots & a_n^{k-1} \end{bmatrix}$$

多项式码是一批好码，它的缺点是码长太小($n \leqslant q$).

- 所以,期望得分: >=30

算法五_二元Reed-Muller码

我们知道，每个 m 元布尔函数 $f=f(x_0,x_1,\cdots,x_{m-1}):\ \mathbb{F}_2^m\to\mathbb{F}_2$ 可以表示成多项式

$$\begin{aligned} & f(x_0,x_1,\cdots,x_{m-1}) \\ = & \sum_{(a_0,a_1,\cdots,a_{m-1})\in\mathbb{F}_2^m} f(a_0,a_1,\cdots,a_{m-1})(x_0+a_0+1)(x_1+a_1+1)\cdot\cdots\cdot(x_{m-1}+a_{m-1}+1) \\ = & \sum_{i_0,i_1,\cdots,i_{m-1}=0}^{1} C(i_0,i_1,\cdots,i_{m-1})x_0^{i_0}x_1^{i_1}\cdot\cdots\cdot x_{m-1}^{i_{m-1}} \quad (C(i_0,i_1,\cdots,i_{m-1})\in\mathbb{F}_2). \end{aligned}$$

$$c_f=(f(v_0),f(v_1),\cdots,f(v_{n-1}))\in\mathbb{F}_2^n \quad (n=2^m),$$

其中 $v_0,v_1,\cdots,v_{n-1}$ 是以某种固定次序列出的 $\mathbb{F}_2^m$ 中全部 $n=2^m$ 个向量. 比如对每个 $i,0\leqslant i\leqslant n-1=2^m-1$，令

$$i=i_0+i_1\cdot 2+\cdots+i_{m-1}\cdot 2^{m-1}(i_\lambda\in\{0,1\})$$

为 i 的二进制展开，可取 $v_i=(i_0,i_1,\cdots,i_{m-1})$. 于是

$$v_0=(0,0,\cdots,0),v_1=(1,0,\cdots,0),\cdots,v_{2^m-1}=(1,1,\cdots,1).$$

算法五_二元Reed-Muller码

将 m 元布尔函数对应于向量 $c_f \in \mathbb{F}_2^n$，则给出全部 m 元布尔函数组成的集合 B_m 和向量空间 $\mathbb{F}_2^n$ 之间的一一对应

$$\begin{array}{ccc} B_m & \overset{1:1}{\longleftrightarrow} & \mathbb{F}_2^n \quad (n=2^m) \\ f(x_0,x_1,\cdots,x_{m-1}) & \mapsto & c_f=(f(v_0),f(v_1),\cdots,f(v_{n-1})). \end{array}$$

这是 $\mathbb{F}_2$ 向量空间同构：对于 $f,g \in B_m$，有 $c_{f+g}=c_f+c_g$. 并且

$$c_{fg}=(f(v_0)g(v_0),f(v_1)g(v_1),\cdots,f(v_{n-1})g(v_{n-1})).$$

算法五_二元Reed-Muller码

定义 设 $m \geqslant 1, n = 2^m, 0 \leqslant r \leqslant m$. 则 $\mathbb{F}_2^n$ 的子集合

$$RM(r,m) = \{c_f \in \mathbb{F}_2^n \mid f \in B_m, \deg f \leqslant r\}$$

叫作 r 阶二元 R－M 码.

由于 B_m 中满足 $\deg f \leqslant r$ 的所有 m 元布尔函数形成 $\mathbb{F}_2$ 上的向量空间，可知 $RM(r,m)$ 是 $\mathbb{F}_2^n$ 的线性子空间，即 $RM(r,m)$ 是二元线性码. 又次数小于等于 r 的所有单项式所对应的码字

$$\{c_f \mid f = x_{i_1} x_{i_2} \cdots x_{i_t}, 0 \leqslant t \leqslant r, 0 \leqslant i_1 < i_2 < \cdots < i_t \leqslant m-1\}$$

形成二元线性码 $RM(r,m)$ 的一组基. 于是此码的维数为

$$k = k(r,m) = \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{r}.$$

算法五_二元Reed-Muller码

定理 $0 \leqslant r \leqslant m$. 则二元 R－M 码 $RM(r,m)$ 的参数为 $[n,k,d] = [2^m, k(r,m), 2^{m-r}]$，其中 $k(r,m) = \sum_{i=0}^{r} \binom{m}{i}$.

期望得分: >=30

Singleton界

定理 （Singleton 界） 如果存在参数(n,K,d)的 q 元码，$1 \leqslant d \leqslant n-1$. 则 $K \leqslant q^{n-d+1}$（即 $n \geqslant k+d-1$）.

定义 满足 Singleton 界的码（即 q 元码(n,K,d)，$K=q^{n-d+1}$），叫做极大距离可分码，简称 MDS 码（Maximal Distance Separable）.

这也是一类较好的纠错码.

Question time

Thank you